DOCUMENT RESUME

ED 292 341 FL 017 258

AUTHOR Jackson, T. Wendell; Clegg, J. Halvor
TITLE Strategies for the Teaching of Vocabulary Based on Oral Frequency Counts.
PUB DATE 88
NOTE 11p.; In: Lalande, John F. II, Ed., Shaping the Future of Language Education: FLES, Articulation, and Proficiency. Report of Central States Conference on the Teaching of Foreign Language; see FL 017 249.
PUB TYPE Information Analyses (070) -- Guides - Classroom Use - Guides (For Teachers) (052) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Classroom Techniques; Computer Oriented Programs; Dialects; Form Classes (Languages); Language Research; Language Usage; *Oral Language; Second Language Instruction; *Spanish; Verbs; *Vocabulary Development; *Word Frequency

ABSTRACT

Previous research on Spanish vocabulary for second language instruction has focused primarily on word frequency in a variety of print sources, and has the added disadvantages of being outdated and having too complex a format. A new, computer-based analysis of spoken Spanish from different dialect regions of the world was undertaken to provide more useful information for teaching core vocabulary. The results are similar to those of earlier studies, showing that: there is considerable semantic divergence in personal vocabulary choice; function words comprise the vast majority of high frequency vocabulary; a nucleus of common adverbs provides a basis for general instruction; a nucleus of eight common adjectives also provides a basis for general instruction; the 10 most frequent verbs account for more than half of total usage; four verb tenses account for over 80 percent of forms used; the present tense accounts for over 50 percent of forms used; the infinitive is the second most used verb form; some individual conjugated forms comprise large proportions of overall frequency; some verb forms occur in less than one-half of the one percent of usage; and some verb forms tend to "specialize" in a limited number of forms. The results and their implications for instructional approaches and emphases are discussed. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED292341

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

9 Strategies for the Teaching of Vocabulary Based on Oral Frequency Counts

T. Wendell Jackson
Brigham Young University
J. Halvor Clegg
Brigham Young University

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

L. Fiddle

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The recent attention to the importance of oral proficiency and consequent methodologies for its achievement has helped us to become more aware of the role of vocabulary in language acquisition. Higgs and Clifford state in their Hypothesized Relative Contribution Model: "Vocabulary is obviously more essential than is grammar for successful performance at Level I tasks" (Higgs and Clifford, 6, pp. 68–69). This is significant since the authors clearly consider these two skills to be the more important skills taught, and relegate pronunciation and culture to minimal roles in performance. If we agree with Higgs and Clifford that vocabulary is of primary importance in early language learning, then we must deal with the question of what vocabulary to teach and which strategies are effective in teaching selected items.

The History of Word Frequency Studies in Spanish

FL017258

Modern research on Spanish vocabulary received its first impetus in the 1920s. There was a good deal of concern over what should be included in vocabulary lists since the only ones available were those compiled by textbook authors. Hayward Keniston, one of the leading language specialists of the era, was severely critical of existing lists,

commenting that when an authoritative list did become available, there would "be less excuse for the perverted vocabularies which now mar so many of our Spanish grammars, composition-books, and readers" (Keniston, 8, p. 88).

In 1927 Milton A. Buchanan published his Spanish word count, *A Graded Spanish Word Book*. The purpose of his work was to "provide material for graded vocabulary tests" and was carried out under the auspices of the American and Canadian Committees on Modern Languages (Buchanan, 1, p. 4). He analyzed some 1,200,000 words from among seven categories (plays, novels, verse, folklore, miscellaneous prose, technical literature and periodicals). He included texts from both Latin America and Spain and went back as far as the seventeenth century to include works such as the *Quijote* and the Bible. Buchanan's work became the principal work in the field despite criticism that it represented literary language rather than the common idiom. It remained unchallenged until 1952, when Ismael Rodríguez Bou created his *Recuento de vocabulario español* (9). His desire was to create a work that could be used in the teaching of Spanish and the creation of textbooks. He included some 7,000,000 words from varying sources. He incorporated the Buchanan graded word book, oral vocabulary, school compositions, radio programs, religious literature, textbooks, and more. His primary sources were principally Puerto Rican and collected during the decade from 1942 to 1952.

One year later, in 1953, Víctor García Hoz published *Vocabulario usual, vocabulario común y vocabulario fundamental* (4), a study of the language of the common Spaniard, totaling approximately 400,000 words. His sources included private letters, periodicals, official documents, and the best-selling books of the previous decade.

The last major frequency count published for Spanish appeared in 1964, viz., the *Frequency Dictionary of Spanish Words* (7) of Alphonse Juilland and Eugenio Chang-Rodríguez. Their data were compiled in a much more scientific manner and based on selected Spanish literature published between 1920 and 1940. The areas used were dramatic, fictional, essayistic, technical and journalistic. The total number of words studied was 500,000.

A review of these four frequency lists reveals the widely varying orientations of the authors. There is vocabulary dating to the seventeenth

century in two of the works. The dialectal variety of the last three works is very narrow, representing only one country (either Spain or Puerto Rico). A shared limitation of all four lists is that the near totality of all four lists reflects *written* language. Further, the most recent eras of language studied date back to the early 1960s and now represent the language of parents or grandparents.

Some of the results of this type of research are universal, however. In every case, including our own, the most frequent words are largely the "function" words; i.e., prepositions, conjunctions, articles, and so on, have the largest numbers of occurrences. Buchanan (1, p. 10) discarded these (approximately 189 words) from his study, as did Rodríguez Bou (approximately 105). Rodríguez Bou (9, p. 654) noted that these words comprised 50 percent of the first million words studied.

The only list currently being published is that of Juilland and Chang-Rodríguez (7). In spite of the intent of the authors and in spite of its merit, their frequency dictionary is a difficult source for teachers and text writers to use. Along with the problems of source and era, the format is a major difficulty. All of the inflected forms are grouped under one heading; hence, all forms of *ser* are listed under *ser*. The data for individual forms is presented in an *alphabetized* section giving no indication of the relationship of individual entries to the relative frequency of all the items in the work. The entry *yo*, for example, includes *yo*, *me*, *mí*, *-migo*, *nosotros*, *nosotras*, *nos*, and so on, thus rendering a rather awkward perspective on *yo*. The same is true for *ser*. All of the morphological variants, such as *ser*, *fuimos*, *sido*, *somos*, *es*, are found under *ser*. The form *es* provides 47.7% of the occurrences of *ser*. This figure would place it *sixteenth* on their frequency list, yet *es* does not appear there at all.

Of all these sources, the only one to include spoken materials was Rodríguez Bou (9). Unfortunately, his materials are buried inconveniently in the body of the text. As a result, no spoken materials are readily available to the teacher and text writer of Spanish.

In addition to the serious questions on written sources, outdated time period, and complex format, the research leads to the question of the utility of high frequency forms. In a recent article dealing with Asher's Total Physical Response strategy as a technique for teaching all

skills in Spanish, Glisan proposes that "the author's personal experience has shown that more effective learning results when students know *well* a small core of the most important, high frequency structures" (Glisan, 5, p. 424).

Glisan's position is borne out by the research. In addition to the observation made by Rodríguez Bou (9, p. 654) that his 109 function words comprise some 50 percent of the first million words he studied, Juilland and Chang-Rodríguez indicate that the most common 500 words comprise 79.1 percent of the 500,000 words they studied (Juilland and Chang-Rodríguez, 7, p. xliii).

The Current Study on Oral Frequency

In the absence of research that could provide contemporary data, a broader dialectal base, *spoken* sources, and a "user friendly" format, we initiated a computer-based analysis of spoken Spanish from different dialect regions of the Hispanic world. This was facilitated by the existence of materials created by the Programa Interamericano de Lingüística y Enseñanza de Idiomas (PILEI) (now carried out by the Asociación de Lingüística y Filología de América Latina [ALFAL]) on Spanish spoken in the major cities of the Spanish-speaking world. The data for this research come from materials carried out, and published by, the Centro de Lingüística Hispánica de la Universidad Nacional Autónoma de México (participants in the PILEI project) and from materials concerning Costa Rica that were published by Miguel Salguero in his *Así vivimos los ticos* (10). The Costa Rican materials were added because the PILEI group did not include a city from Central America in its project.

The materials from Mexico are found in *El habla popular de la ciudad de México* (3). These come from interviews conducted by the members of the Centro de Lingüística Hispánica of the Universidad Nacional Autónoma de México. They represent the speech of the educated (*culto*) Mexican. The format included interviews with one person, interviews with two persons, secret interviews, and conference talks. Our samples were collected at random from each section of the book. The second page of an interview was selected and then every seventh

page thereafter until a total of 10,000 words was accumulated. The samples included the speech of nine males and four females. It was not possible to control the sex factor due to the methods of interviewing and sampling.

The Costa Rican materials are also in the form of interviews. The samples were taken with regard to level of education as well. A total of five informants (four males and one female) was included. The number of females was again restricted by the makeup of the book, which had only three interviews with females. Nevertheless, a sample of female speech was available. The sample size was somewhat larger, viz., 15,000 words.

The printed pages were entered into an IBM 370 computer by means of a Kurzweil optical scanner. The data were then extensively edited for correctness. A secondary editing was then conducted to provide tagging for homophones and grammatical categories. A concordance program was written along with an alphabetical and numerical sort routine. The data were then concordanced and sorted. The third (and longest) editing task grouped all of the morphological variants of a basic form, such as all of the variants of *ser*.

The results of this research are similar to those of earlier studies. They show the same high frequency function words: conjunctions, prepositions, articles, numbers, and pronouns. These forms vastly outnumber the other word forms in frequency of occurrence—despite their limited number. They include forms unique to a dialect, such as the Mexican *pues* in all its varieties. A review of the Juilland and Chang-Rodríguez (7) word frequency count shows a total of 19 conjunctions, 15 prepositions, 6 articles, 35 numbers, and 52 pronouns (a total of 127 forms), out of 5024 words. The fact that half of Juilland and Chang-Rodríguez's 500,000 occurrences fall among the 36 most frequent forms (this includes 8 high frequency verbs, 2 adverbs, and 3 adjectives) makes the importance of these function words even greater.

The other word forms (adverbs, verbs, nouns, and adjectives) have lower rates of occurrence, but with the exception of the adverbs, have much higher numbers of forms. Juilland and Chang-Rodríguez includes 185 adverbs, 957 verbs, 2549 nouns (just over half of the total), and 1199 adjectives. The same general frequency pattern holds for these materials.

One area of noteworthy difference is in the higher frequency of occurrence of interjections in the oral materials. Juilland and Chang-Rodríguez list no interjections in the first 500 words while in the Mexican oral materials word number 52 was the interjection *ah*, number 83 was *eh*, number 139 was *ay*, and number 191 was *m*.

One of the important results was the semantic divergence among the various sources. The function words were generally the same in all sources. The nouns, verbs, adjectives, and adverbs were generally diverse. This was especially the case with the nouns.

A review of the first 250 words from each source provides a total of 106 nouns. The Mexican corpus includes 20 nouns (10 of the 20 were unique to the Mexican corpus). The Costa Rican corpus had 49 nouns (29 of the 49 were unique). Juilland had 66 nouns (44 of the 66 were unique). There were only *six* nouns shared among all three sources. These nouns were: *vez*, *cosa*, *año*, *casa*, *campo*, and *trabajo*. In addition, among the 106 nouns there were only 23 duplications (including the six shared nouns).

The relationship among the adverbs was closer. The total number of adverbs used in the Spanish language is very limited (Juilland and Chang-Rodríguez list 185). Juilland and Chang-Rodríguez give 38 adverbs (with 17 unique words) in their first 250 words. The Mexican corpus includes 20 (with *ahí* and *abajo* being unique to it). The Costa Rican corpus has only 13 (*cómo*, *sólo*, and *actualmente* being unique). There are *10* shared adverbs among the three sources and 18 duplications (including the 10 shared adverbs). The shared adverbs were: *no*, *más*, *sí*, *ya*, *muy*, *así*, *aquí*, *también*, *bien*, and *dónde*.

The adjectives show a different correlation. There is a higher frequency nucleus of 8 shared adjectives and then a considerable degree of divergence. The nucleus includes the following: *este*, *bueno*, *ese*, *mucho*, *otro*, *todo*, *cuanto*, and *grande*. The total number among the first 250 words was 41 adjectives. The Mexican corpus had 20 (including the unique words *chico*, *feo*, and the expletive *pinche*). The Costa Rican corpus had 24 (with 6 unique forms). Juilland had 31 (with 13 unique forms). The total number of shared adjectives was 13. The number of duplications was 19 including the shared adjectives.

Verb usage in spoken Spanish in Costa Rica and Mexico demonstrates that there is a relatively small number of high frequency verbs

that account for a large percentage of total usage. The first 10 of more than 200 verbs account for more than one half (50.3 percent) of the total count. These verbs, their frequencies, and average percentage of use are shown in Table 1.

Table 1
High Frequency Verb Forms (Oral)

Freq.	*Verb*	*Inf.*	*Pres. Ind.*	*Pret.*	*Imp.*	*Pres. Subj.*	*Past Subj.*	*Past Part.*	*Cond.*
1343	Ser	6	69	8	7	2	6	1	1
761	Haber	2	81	1	9	3	3	1	1
563	Tener	8	65	5	10	3	1	4	1
452	Decir	12	57	17	4	4	1	3	2
424	Ir	15	62	4	9	4	0	4	2
419	Estar	4	66	12	12	2	0	4	0
286	Hacer	25	52	10	2	3	3	5	0
243	Ver	39	45	3	3	5	0	5	0
223	Poder	9	57	4	4	5	2	7	11
199	Dar	18	50	15	6	6	3	1	2
Average use (%)		14	60	7	7	4	2	4	2

The research also indicates that only a few forms of these selected high frequency verbs account for the vast majority of usage. We have learned, for example, that only four verb forms, i.e., infinitive, present indicative, preterit, and imperfect account for 88 percent of the verbs given in the table. This is also generally true for the entire set of verbs analyzed. The frequency of present indicative forms is overwhelming, indicating that 60 percent of all forms used are in the present tense. Another meaningful finding is in the number of infinitives that are used. The 14 percent frequency of infinitives is as high as the preterit and imperfect tenses combined. The relative importance of the infinite is largely ignored among text writers and grammarians. On the other hand, the future and imperatives were so low in frequency that they were not significant in the overall results. By the same token, the past subjunctive, the conditional, and the progressives were barely used. The perfect tenses and the present subjunctive were used slightly more, but of all these forms none reached 10 percent.

Furthermore, some individual conjugated forms dominate the frequency of these verbs. For example, the form *es* made up 246 of 564 total occurrences of the verb *ser* in the Mexican materials, a total of 43.6

percent. The form *está* made up 105 of 272 total occurrences of the verb *estar*, a total of 38.6 percent. The form *hay* made up 62 of 190 total occurrences of the verb *haber*, a total of 32.6 percent. These numbers, by themselves would place these three individual forms among the top 25 verbs in frequency.

Beyond the ten most frequent verbs the divergence becomes much more apparent. In the first 250 words for all sources there appear 70 different verbs. Fifteen of these verbs are shared among all sources; 30 of the verbs are duplicated (including the shared verbs). Forty of the 70 verbs appear only in one source.

What the above analysis for verbs would seem to indicate is that students must have extensive practice with selected verbs and particularly in the four forms more frequently used, as well as in certain high-frequency conjugated forms, such as *es*, *está* and *hay*. Thus, if students concentrate their efforts on high proficiency with such selected forms, their ability to communicate in the target language, as well as their consequent motivation, should be significantly enhanced. The results also illustrate the significance of the need to identify such high frequency words as a tool in designing curriculum.

Other verbs tend to "specialize" in the forms they more commonly use, such as *fijar*, *mirar*, and *echar* were most frequently used in the imperative (*fijate*, *mira*, and *échale*).

Conclusions

With the attention given to vocabulary in current foreign language pedagogy, a review of available materials shows a lack of vocabulary materials that are contemporary, of a broad dialectal base, based on *spoken* language, and readily usable. The results of our analysis of such materials lead to the following conclusions:

1. There is considerable semantic divergence in personal vocabulary choice. This is particularly the case with nouns.
2. Function words comprise the vast majority of high frequency vocabulary.
3. A nucleus of common adverbs provides a basis for general instruction.

4. A nucleus of eight common adjectives also provides a basis for general instruction.
5. The most frequent ten verbs in both oral and written materials account for more than one half of total usage.
6. The infinitive, present indicative, preterit, and imperfect account for over 80 percent of forms used.
7. The present tense accounts for over 50 percent of forms used.
8. The infinitive is the second most used verb form.
9. Some individual conjugated forms, such as *es*, *está* and *hay* comprise large portions of the overall frequency.
10. Some forms, such as the future and the imperatives, are so little used that they occur less than one half of 1 percent.
11. Some verbs tend to "specialize" in a limited number of forms, such as *fijar* in *fijate*, and so on.

Recommendations

The following recommendations for instruction in Spanish are supported by our study on oral frequency of vocabulary.

1. That oral vocabulary be emphasized or treated separately in instruction
2. That the high frequency function words be emphasized
3. That semantic divergence be accounted for, particularly with regard to nouns
4. That a nucleus of common adverbs be taught. These include the adverbs *no*, *más*, *sí*, *ya*, *muy*, *así*, *aquí*, *también*, *bien*, and *dónde*.
5. That a nucleus of common adjectives be taught. These include the adjectives *este*, *bueno*, *ese*, *mucho*, *otro*, *todo*, *cuanto*, and *grande*.
6. That a small core of the most frequent ten verbs be taught thoroughly. These are *ser*, *haber*, *tener*, *decir*, *ir*, *estar*, *hacer*, *ver*, *poder*, and *dar*.
7. That emphasis be placed primarily on the present indicative tense in initial instruction
8. That the present indicative be followed by instruction in the uses of the infinitive

9. That the imperfect and preterit tenses be the only other tenses taught initially
10. That high-frequency conjugated forms, such as *es*, *está*, and *hay* be given emphasis in early instruction
11. That forms such as the future, the imperative, the subjunctives, the conditional, the progressives, and the perfects be taught only in fixed phrases initially
12. That more frequent oral forms be included in materials
13. That specialized verbs be emphasized only in their frequent forms

References

1. Buchanan, Milton A. *A Graded Spanish Workbook*. Toronto: University of Toronto Press, 1927.
2. Centro de Lingüística Hispánica. *El habla de la ciudad de México*. México: Universidad Nacional Autónoma de México, 1976.
3. Centro de Lingüística Hispánica. *El habla popular de la ciudad de México*. México: Universidad Nacional Autónoma de México, 1976.
4. García Hoz, Víctor. *Vocabulario usual, vocabulario común y vocabulario fundamental*. Madrid: Gráficas Orbe, 1953.
5. Glisan, Eileen W. "Total Physical Response: A Technique for Teaching All Skills in Spanish." *Foreign Language Annals* 19 (1986): 419–27.
6. Higgs, Theodore V., and Ray Clifford, "The Push toward Communication," pp. 57–79 in Theodore V. Higgs., ed., *Curriculum, Competence, and the Foreign Language Teacher*. The ACTFL Foreign Language Education Series, vol. 13. Lincolnwood, IL: National Textbook Co., 1982.
7. Juilland, Alphonse, and Eugenio Chang-Rodríguez. *Frequency Dictionary of Spanish Words*. The Hague: Mouton & Co., 1964.
8. Keniston, Hayward. "Common Words in Spanish." *Hispania* 3 (1920): 85–96.
9. Rodríguez Bou, Ismael. *Recuento de vocabulario español*. Baltimore, MD: Waverly Press, 1952.
10. Salguero, Miguel. *Así vivimos los ticos*. San José, Costa Rica: EDUCA, 1976.